DEMOCRATIZING AND EXPANDING THE REACH OF DIA MASS SPECTROMETRY: DEVELOPING OPENSWATH TOOLS AND WORKFLOWS WITHIN USER-FRIENDLY GALAXY-P PLATFORM Pratik Jagtap^{1,2}; Sarah Parker³; Bjoern Gruening⁴; Ira Cooke⁵; James Johnson⁶, Hannes Roest⁷, Ceorge Rosenberger⁷, Tzu-Yi Yang², Candace Guerrero², Brett Noel²; Laurie Parker², Jennifer Van Eyk³, Ruedi Aebersold^{7, 8} and Tim Griffin^{1,2}

¹ Center for Mass Spectrometry and Proteomics, UMN, St. Paul, MN; ² Biochem. Mol Biol. Biosystems Research Institute, Los Angeles, CA; ⁴ University of Freiburg, Freiburg, Germany; ⁵La Trobe University, Melbourne, Australia; ⁶Minnesota Supercomputing Institute, UMN, Minneapolis, MN; ⁷ Department of Biology, IMSB, ETH Zurich, Zurich, Zurich, Switzerland; ⁸Faculty of Science, University of Zurich, Zurich, Switzerland



used.

benSwathWorkflow Complete workflow to run OpenSWATH (Galaxy Tool Version 2.0.0) Options t files separated by blank 34: OpenSwathWorkflow on data 4, data 3, and others Image: Separate Sep	DATA CONVERSION	WRAP	INSTALL	TEST	
32: OpenSwathWorkflow on data 4, data 3, and others 1: SJP_140527_ECB_SWATH-64v-SJP_006.mzML.	Define .wiff datatype				
sition file ('TraML','tsv' or 'csv')	msconvert to convert .wiff files to mzml				
put file type default: determined from file extension or content ansition file ('TraML')	MS data Converter from AB Sciex				U.
Image: Comparison of the second se	SEARCH				
output file (mProphet compatible) extract RT around this value (-1 means extract over the whole range, a value of 600 means to extract around +/-	msconvert (ProteoWizard)				
s of the expected elution) 0 extraction_window)	indexmzXML (TPP) X! tandem				
ctraction window used (in Thomson, to use ppm see -ppm flag) /z extraction_window is in ppm reporting after feature (ordered by quality; -1 means do not stop)	Comet PeptideProphet Xinteract (TPP)				
p_report_after_feature)	idconvert (TPP)				
normalized RT is expected to be between 0 and 1. If your normalized RT has a different range, pass this here (0 normalization_factor) e.g. it goes from 0 to 100, set this value to 100)	InterProphetParser (TPP)				
finding after feature (ordered by intensity; -1 means do not stop) p. after feature)	PeptideShaker				
mal peak width (s), discard all peaks below this value (-1 means no action)	LIBRARY BUILD				
n_peak_width) s to get better peak picking by looking at peak consistency of all picked peaks	Mayu.pl (TPP)				c
alculate_peaks) Tries to use the consensus (median) peak border if theof variation within the picked peaks is too large rmines the maximal Z-Score (difference measured in standard deviations) that is considered too large for peak idaries	SpectraST (TPP)				
'S alculate_peaks_max_z) If the Z-Score is above this value, the median is used for peak boundaries (default value 1.0) if compute_peak_quality is set, this parameter will not consider peaks below this quality threshold	Spectrast2spectrast_iRT.py				
.5 nimal_quality) s to compute a quality value for each peakgroup and detect outlier transitions	LIBRARY REFINEMENT		3 <u></u>		
s to compute a quality value for each peakgroup and detect outlier transitions mpute_peak_quality) The resulting score is centered around zero and values above 0 are generally good and below -1 or -2 sually bad	SpectraST (TPP)				
number of subsequent data points used for smoothing	Spectrast2tsv.py TSVtoTraML				
play_frame_length) number has to be uneven. If it is not, 1 will be added er of the polynomial that is fitted	OpenSWATHDecoyGenerator				
play_polynomial_order) sian width in seconds, estimated peak size	OpenSWATH WORKFLOW				6
uss_width) Gaussian filter for smoothing (alternative is Savitzky-Golay filter)	Main OpenSWATH_Workflow (OpenMS)				G
e _gauss) e a certain minimal peak_width on the data (pyprophet				G
ak_width) e.g. extend the peak at least by this amount on both sides) in seconds1 turns this feature off gnal-to-noise threshold at which a peak will not be extended any more. Note that setting this too high (Feature aligner				
ry to remove overlapping peaks during peak picking hich method to choose for chromatographic peak-picking (OpenSWATH legacy, corrected picking or Crawdad)	STATISTICAL ANALYSIS				
IA extraction window in Th se centroded DIA data IA b/y series minimum intensity to conside	R MSStats Scripts				(
IA nr of isotopes to conside IA nr of charges to conside					
IA maximal difference in ppm to count a peak at lower m/z when searching for evidence that a peak might not be oisotopic mum number of iterations using by Levenberg-Marquardt algorithm	CURREN	T STATU	US		 •
x_iteration)		00.001			
taRelError) anced Options	 Tools such as wi 	-	-		
w Advanced Options v RT normalization file (how to map the RTs of this run to the ones stored in the library) tional table constraints file containing the SWATH windows leave affect years affect years and a stored in the library	msconvert (for v	viff files),	Sciex I	MS	
tional, tab separated file containing the SWATH windows: lower_offset upper_offset \newline 400 425 \newline 4: swaths_64_header.txt wath_windows_file) Note that the first line is a header and will be skipped	Data Converter,	OpenSW	ATH		•
Sort of input SWATH files when matching to SWATH windows from swath_windows_file Extract the precursor ion trace(s) and use for scoring nimal distance to the edge to still consider a precursor, in Thomson	Workflow and F	vprophet	t have l	been	
) nin_upper_edge_dist)	installed and tes	• • •	· • •		
Output an XIC with a RT-window that by this much larger (nimum r-squared of RT peptides regression .95	 Tools within Tra 		mio		
nin_rsq) nimum relative amount of RT peptides to keep 0.6		—			
nin_coverage) The input files each contain one single SWATH (alternatively: all SWATH are in separate files)	pipeline (TPP) k		install	ea	
Turn on elution model score (EMC fit to peak) Whether to run OpenSWATH directly on the input data, cache data to disk first or to perform a datareduction step st	but not tested ye	et.			
ache adOptions) If you choose cache, make sure to also set tempDirectory generaty directory to store cached files for example	• The following to	ools have	been		•
tmp/ empDirectory) Function used to extract the signal	wrapped and ar	e on the g	galaxy		
The batch size of chromatograms to process (0 means to only have one batch, sensible values are around 500- 00) Overwrite tool specific checks Cutoff in m/z below which peaks should not be used for quantification any more	toolshed and or	· · · ·	- •		
Cutoff in m/z below which peaks should not be used for quantification any more Whether to write out all points of all features into the featureXML Use the shape score (this score measures the similarity in shape of the transitions using a cross-correlation)	- SpectraST to T	0			
Use the coelution score (this score measures the similarity in coelution of the transitions using a cross-correlation) Use the retention time score (this score measure the difference in retention time) Use the library score	1		anhat a	nd	
Use the intensity score Use the number of peaks score Use the total XIC score	- Peptide Prophe	· 1	1	ſΙ	
Use the SN (signal to noise) score Use the DIA (SWATH) scores Use the correlation scores with the MS1 elution profiles	Protein Prophet from	n the TPF	•		
Use the full MS1 scan at the peak apex for scoring (ppm accuracy of precursor and isotopic pattern) Execute	- OpenMS tools.				
	- Feature Aligner				a



Link: https://github.com/galaxyproteomics toolshed: https://toolshed.g2.bx.psu.edu/repository google group: z.umn.edu/galaxypuserforum

FUTURE PLANS

- an to continue on our implementation of OpenSWAT ools in user-friendly Galaxy
- work to increase usability, dissemination and enable ation with multi-omic approaches.
- an to develop workflows that will offer user-friendly rsion and analysis of file formats.
- that offer an alternative to Transproteomic Pipeline ts such as SearchGUI / PeptideShaker will also be oped and tested for integration into OpenSWATH low.
- stream tools such as MSStats for statistical
- itative analysis of selected peptides will also be ated and tested.
- native DIA workflows such as DIA-Umpire that use r tools will also be explored.



xyP is supported through the NSF Grant # 1147079.